

Comparison of visual vs. automated detection of lipemic, icteric and hemolyzed specimens: can we rely on a human eye?

Ana-Maria Simundic*, Nora Nikolac, Valentina Ivankovic, Dragica Ferenec-Ruzic, Bojana Magdic, Marina Kvaternik and Elizabeta Topic

University Department of Chemistry, School of Medicine, Faculty of Pharmacy and Biochemistry, Zagreb University, University Hospital "Sestre Milosrdnice", Zagreb, Croatia

Abstract

Background: Results from hemolyzed, icteric, and lipemic samples may be inaccurate and can lead to medical errors. These preanalytical interferences may be detected using visual or automated assessment. Visual inspection is time consuming, highly subjective and not standardized. Our aim was to assess the comparability of automated spectrophotometric detection and visual inspection of lipemic, icteric and hemolyzed samples.

Methods: This study was performed on 1727 routine biochemistry serum samples. Automated detection was performed using the Olympus AU2700 analyzer. We assessed: 1) comparability of visual and automated detection of lipemic, icteric and hemolyzed samples, 2) precision of automated detection, and 3) inter-observer variability for visual inspection.

Results: Weighted κ coefficients for comparability of visual and automated detection were: 0.555, 0.529 and 0.638, for lipemic, icteric and hemolyzed samples, respectively. The precision for automated detection was high for all interferences, with the exception of samples being only slightly lipemic. The best overall agreement between observers was present in assessing lipemia (mean weighted $\kappa=0.698$), whereas the lowest degree of agreement was observed in assessing icterus (mean weighted $\kappa=0.476$).

Conclusions: Visual inspection of lipemic, icteric and hemolyzed samples is highly unreliable and should be replaced by automated systems that report serum indices.

Clin Chem Lab Med 2009;47:1361–5.

Keywords: hemolysis; ictericia; interference; lipemia; preanalytical.

Introduction

Most errors occur in the preanalytical phase of clinical laboratory testing (1, 2). Successful monitoring and management of preanalytical sources of interferences is therefore crucial to the quality of laboratory diagnostic process and to the quality of patient care. Results from hemolyzed, icteric, and lipemic samples may be inaccurate and can lead to medical errors, and as such represent a considerable hazard to patient health (3, 4). Modern clinical chemistry analyzers are equipped with automated systems for detection of lipemic, icteric, and hemolyzed samples. With continuous technological development and laboratory automation, a significant reduction in laboratory errors can be expected (5, 6). However, in laboratories that do not use automated systems for detection and management of preanalytical interferences, unsuitable samples are detected by means of visual inspection by individual laboratory personnel. Visual inspection is not only time consuming, but also highly subjective, non-standardized and may be a potential source of error (7, 8). Our goal was to assess the comparability of automated spectrophotometric detection and visual inspection of lipemic, icteric and hemolyzed samples. Automated detection was performed using the Olympus AU2700 (Olympus, Tokyo, Japan) clinical chemistry analyzer with the manufacturer's reagents. To the best of our knowledge, this type of study has not yet been performed.

Materials and methods

The University Department of Chemistry is an ISO 15189 accredited central laboratory at the tertiary care 900 bed University Hospital Sestre Milosrdnice, in Zagreb, Croatia. The laboratory has a capacity of up to 3 million tests per year. All comparisons were performed in our laboratory during July 2008. We assessed the following: 1) comparability of visual and automated detection of lipemic, icteric and hemolyzed samples, 2) precision of the Olympus LIH reagent (Olympus, O'Callaghan's Mills, Co. Clare, Ireland), and 3) inter-observer variability for visual inspection of samples.

Comparability of visual and automated detection

The comparability assessment was performed using 1727 consecutive routine biochemistry serum samples. We used routine in-patient hospital serum samples collected outside the laboratory by the clinical staff. Blood was obtained by venipuncture and collected in evacuated BD Vacutainer® glass tubes without any additive (Becton-Dickinson, Franklin Lakes, NJ, USA). Samples were centrifuged at 3500 g for 10 min. After centrifugation, serum was inspected visually

*Corresponding author: Assist. Prof. Ana-Maria Simundic, PhD, University Department of Chemistry, Department of Molecular Diagnostics, University Hospital "Sestre Milosrdnice", Zagreb, Croatia
E-mail: am.simundic@gmail.com
Received June 30, 2009; accepted July 17, 2009;
previously published online September 25, 2009

Table 1 Degrees of interference assessed visually and by use of the LIH Olympus reagent, and the corresponding concentration of the interfering substance.

Interference degree	0	1	2
Lipemia			
Visual	No	Mild	Severe
LIH	N	+	(2+)–(5+)
Intralipid, g/L ^a	<0.4	0.4–1	>1
Icteria			
Visual	No	Mild	Severe
LIH	N	(+)–(2+)	(3+)–(5+)
Bilirubin, $\mu\text{mol/L}$ ^a	<42.7 $\mu\text{mol/L}$	42.7–169.3 $\mu\text{mol/L}$	>169.3 $\mu\text{mol/L}$
Hemolysis			
Visual	No	Mild	Severe
LIH	N	+	(2+)–(5+)
Hemoglobin, g/L ^a	<0.5 g/L	0.5–1 g/L	>1 g/L

^aAccording to the manufacturer (Olympus).

by one observer who was informed about the study. Afterwards, the samples were processed for automated detection using the LIH reagent. The LIH reagent uses a semiquantitative spectrophotometric measurement and grades interfering substances into one of six categories: normal (N), +, 2+, 3+, 4+ and 5+. The spectrophotometric measurement is performed at the following wavelengths (nm): 660 and 800 for lipemia, 480, 570, 600 and 800 for icterus and 410, 480, 600 and 800 for hemolysis. Visual inspection was performed according to our written standard operating procedure for the detection and management of unsuitable specimens. The visual procedure uses standardized colored diagrams with different intensities of lipemia, icterus and hemolysis with the respective concentration of the interfering substance. Interference grading was done according to Table 1.

Precision of Olympus LIH reagent

Within-run precision was determined using patient samples measured 20 times in series. Two patient samples were used for each interference with following concentrations: 8.56 mmol/L and 38.8 mmol/L triglycerides for lipemia, 113.7 $\mu\text{mol/L}$ and 317.6 $\mu\text{mol/L}$ bilirubin for icterus and <1 g/L and 3.5 g/L hemoglobin for hemolysis.

Inter-observer variability for visual inspection

Inter-observer variability of laboratory personnel for visual inspection was examined using a series of 25 serum samples. We compared the results obtained by five different randomly selected laboratory technicians with the individual who visually inspected all samples for the purpose of this study. All observers were informed about the purpose of the procedure. κ coefficients with respective 95% confidence intervals (CI) were calculated for interference detected by each observer.

Statistical analysis

Data were presented as numbers (counts) and percentages. Inter-observer agreement was assessed using a weighted κ coefficient. Interpretation of the κ coefficient was as follows: <0: less than chance agreement; 0.01–0.20: slight agreement; 0.21–0.40: fair agreement; 0.41–0.60: moderate agreement; 0.61–0.80: substantial agreement; 0.81–0.99: almost perfect agreement (9). Precision was expressed as the coefficient of variation (CV%). Statistical analysis was performed using MedCalc[®] statistical software (MedCalc 10.1.3.1, Frank Schoonjans, Mariakerke, Belgium).

Results

Out of 1727 samples, 23 (23/1727 = 1.3%) were flagged as ABN (abnormal) by the Olympus reagent (Table 2). According to the manufacturer's instructions, samples flagged as ABN need to be re-run and/or visually examined. We excluded these samples from further analysis.

Comparability of visual and automated detection

A total of 57 lipemic samples were identified by visual inspection, while automated LIH detection found 35 lipemic samples only (Table 3). The weighted κ coefficient was 0.555 (95% CI = 0.419–0.691) indicating moderate overall agreement.

A total of 101 icteric samples were identified by visual inspection, while automated LIH detection revealed only 74 icteric samples. Visual inspection and automated LIH detection identified a similar proportion of mild and severely icteric samples (Table 4). The weighted κ coefficient was 0.529 (95% CI = 0.427–0.630), showing moderate agreement.

The best comparability was achieved when assessing the grade of hemolysis. Visual inspection and automated LIH detection showed a comparable rate of detection for hemolyzed samples (Table 5). The weighted κ coefficient was 0.638 (95% CI = 0.569–0.707), showing substantial agreement.

Precision of Olympus LIH reagent

Within-run precision (CV) of the LIH reagent was as follows: 0% and 55% at 38.8 and 8.56 mmol/L triglyc-

Table 2 Number of samples flagged by the LIH Olympus reagent as ABN (excluded from further analysis) with different degrees of interference (assessed by visual inspection).

	Visual inspection		
	0	1	2
Lipemia, n	19	4	–
Icteria, n	20	2	1
Hemolysis, n	10	10	3

ABN, abnormal.

Table 3 Comparison of visual inspection and LIH Olympus reagent for lipemia.

LIH Olympus reagent	Visual inspection			
	0	1	2	
0	1642	27	0	1669 (97.9%)
1	5	27	7	33 (1.9%)
2	0	1	1	2 (0.1%)
	1647 (96.7%)	49 (2.9%)	8 (0.5%)	1704

Table 4 Comparison of visual inspection and LIH Olympus reagent for icterus.

LIH Olympus reagent	Visual inspection			
	0	1	2	
0 (n)	1586	41	3	1630 (95.7%)
1 (+ and ++)	17	31	11	59 (3.5%)
2 (3+ to 5+)	0	7	8	15 (0.9%)
	1603 (94.1%)	79 (4.6%)	22 (1.3%)	1704

Table 5 Comparison of visual inspection and LIH Olympus reagent for hemolysis.

LIH Olympus reagent	Visual inspection			
	0	1	2	
0	1529	26	1	1556 (91.3%)
1	27	59	18	104 (6.1%)
2	2	25	17	44 (2.6%)
	1558 (91.4%)	110 (6.5%)	36 (2.1%)	1704

erides for lipemia, 0% and 0% at 317.6 and 113.7 $\mu\text{mol/L}$ bilirubin for icterus and 0% and 10% at 3.5 and <1 g/L hemoglobin for hemolysis. All samples that were not classified correctly were flagged as ABN.

Inter-observer variability for visual inspection

Data on inter-observer variability are shown in Table 6. Five randomly selected laboratory technicians were marked as B-F. The reference technician was marked as A, and all κ coefficients (with respective 95% CI) were calculated in comparison to technician A. The best overall agreement between observers was in the assessment of lipemia. The mean weighted κ value for lipemia was 0.698 (95% CI = 0.626–0.769). The lowest degree of agreement was observed in assessing icterus (weighted κ = 0.476, 95% CI = 0.237–0.716).

Discussion

Unsuitable specimens may significantly affect the accuracy of many clinical chemistry and hematology analyses (10, 11). Successful detection and management of such specimens is one of the major tasks in medical laboratories and an obligation and responsibility for every laboratory accredited according to ISO 15189 standard (12, 13). The prevalence of hemolyzed specimens might be a suitable indicator of preanalytical quality. Hemolysis is most often caused by inadequate specimen collection or sample handling, and occurs with varying prevalence in different health care settings (14). The aim of this study was to explore the agreement between visual and automated detection of lipemic, icteric and hemolyzed samples. Published data on this issue are not common. The key finding of this work is that visual inspection is inferior

Table 6 κ Coefficients with respective 95% CI for five randomly selected laboratory technicians.

	B κ (95% CI)	C κ (95% CI)	D κ (95% CI)	E κ (95% CI)	F κ (95% CI)	Mean κ
Haemolysis	0.584 (0.295–0.874)	0.726 (0.480–0.972)	0.622 (0.318–0.916)	0.569 (0.268–0.870)	0.583 (0.293–0.874)	0.617 (0.537–0.696)
Lipemia	0.743 (0.470–1.016)	0.643 (0.322–0.964)	0.743 (0.470–1.016)	0.627 (0.292–0.962)	0.732 (0.448–1.016)	0.698 (0.626–0.769)
Icteria	0.655 (0.345–0.965)	0.594 (0.276–0.912)	0.160 (–0.321–0.640)	0.446 (0.098–0.794)	0.527 (0.197–0.857)	0.476 (0.237–0.716)
Mean κ	0.661 (0.463–0.859)	0.654 (0.489–0.820)	0.508 (–0.256–1.273)	0.547 (0.318–0.777)	0.614 (0.351–0.877)	

compared with automated detection of preanalytical interferences, such as lipemia, icterus and hemolysis. Poor inter-rater agreement between our observers provides evidence that the human eye is not able to accurately and reproducibly detect subtle differences in sample color and turbidity, even when a standard operating procedure and colored scale are available.

Given that only moderate overall agreement exists between visual and automated detection, and the poor inter-observer agreement, laboratories should not be confident relying on manual visual detection of lipemic, icteric and hemolyzed specimens. Automated systems should be employed and there should be a standard protocol for managing unsuitable samples. Laboratory staff should be continually trained and educated about the importance of the preanalytical quality of laboratory analysis.

The automated detection using the LIH reagent with the AU2700 Olympus analyzer provides precise observations for almost all interfering substances, with the exception of mild lipemia where the LIH showed rather high imprecision. In addition, only 1.3% of samples were flagged as abnormal and required re-assessment. In these instances, visual inspection is irreplaceable.

Lippi et al. compared the efficiency of different analytical platforms for identifying hemolyzed specimens (15). They found that overall imprecision was satisfactory. However, the authors conclude that more effort should be invested in standardizing the way that different laboratories report hemolysis (15).

While there are several reports on the performance of visual and automated detection of hemolyzed samples, information on the agreement of visual assessment and spectrophotometric detection of lipemia and icterus is lacking, as is data on inter-observer agreement. Our results are most consistent with those reported previously by other investigators. Glick et al. examined the accuracy of visual detection of turbidity, hemolysis and icterus in comparison to the measurement triglyceride, hemoglobin and bilirubin concentrations (8). The authors used full color photographs with a scale of six possible grades of interference (trace, 1+, 2+, 3+, 4+ and 5+) for visual detection. Their conclusion was that visual estimation was unreliable. However, it should be noted that sample turbidity only weakly correlates with triglyceride concentration (16), making interpretation of the results by Glick et al. on the accuracy of visual detection of turbidity somewhat questionable.

Another study performed by Hawkins assessed the discrepancy between visual grading and automated hemolysis measurements on the Roche 917 clinical chemistry analyzer (7). Technical staff performing the visual assessment of samples was not informed about the investigation being carried out. For purposes of the study, comments on 800 serum general chemistry results and 800 plasma troponin I results were collected and consolidated with the H index. Hemolysis was overestimated using visual assessment of serum samples and underestimated in plasma samples. The conclusion of this study was that

visual assessment of sample hemolysis is unreliable. Since laboratory personnel was not informed about the study, this observation however does not provide evidence for agreement between visual grading and spectrophotometric determination of the H index measurement. Instead it illustrates the general attitude of the staff towards the importance of hemolysis. Lipemia and icterus interference was not analyzed in this study.

Jeffery et al. compared detection of hemolysis in adult and neonatal samples by inspection and measurement of H index (17). They found that the presence of icterus in neonatal samples decreases the ability to detect hemolysis by visual inspection. Based on their results, the authors conclude that the use of the automated hemolysis index is highly recommended and that potassium in neonatal and adult specimens should be reported with a correction formula, since it might be beneficial to the clinical management of the patient.

One important aspect should also be stressed, namely, the progressive automation and integration of various segments of the total testing process. Such integration of preanalytical modules and analytical workstations would make it difficult to check the specimens before processing. This is why systematic (i.e., automated) inspection for interferences is highly advisable.

Another important issue not covered in our study is the appropriate action taken by the laboratory personnel when interference is detected. There is an ongoing debate as to whether such result should be reported with a comment, corrected or withheld by the laboratory (18). Unfortunately, there is little evidence to support interpretation of laboratory test results and interpretative comments accompanying laboratory reports as a contribution to the outcome of the patient and the quality of the patient care (19, 20). Large well-designed studies assessing the outcome of the patient are needed to define the most appropriate actions to be taken in these instances.

Given the importance of preanalytical interferences as one of the most common sources of laboratory errors, we conclude that visual inspection of lipemic, icteric and hemolyzed samples is highly unreliable and should be replaced by automated systems reporting serum indices. Future studies should define simple algorithms for mathematical correction of test results for interference of hemoglobin, bilirubin and lipids (21).

Acknowledgements

This work is supported by the Ministry of science, education and sports, Republic of Croatia. Project number: 134-1340227-0200.

References

1. Plebani M. Errors in laboratory medicine and patient safety: the road ahead. *Clin Chem Lab Med* 2007;45: 700–7.

2. Plebani M. Laboratory errors: how to improve pre- and post-analytical phases? *Biochimica Medica* 2007;17:5–9.
3. Ryder KW, Glick MR. Erroneous laboratory results from hemolyzed, icteric, and lipemic specimens. *Clin Chem* 1993;39:175–6.
4. Lippi G. Governance of preanalytical variability: traveling the right path to the bright side of the moon? *Clin Chim Acta* 2009;404:32–6.
5. Da Rin G. Pre-analytical workstations: a tool for reducing laboratory errors. *Clin Chim Acta* 2009;404:68–74.
6. Lippi G, Guidi GC. Risk management in the preanalytical phase of laboratory testing. *Clin Chem Lab Med* 2007;45:720–7.
7. Hawkins R. Discrepancy between visual and spectrophotometric assessment of sample haemolysis. *Ann Clin Biochem* 2002;39:521–2.
8. Glick MR, Ryder KW, Glick SJ, Woods JR. Unreliable visual estimation of the incidence and amount of turbidity, hemolysis, and icterus in serum from hospitalized patients. *Clin Chem* 1989;35:837–9.
9. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37:360–3.
10. Lippi G, Blanckaert N, Bonini P, Green S, Kitchen S, Palicka V, et al. Haemolysis: an overview of the leading cause of unsuitable specimens in clinical laboratories. *Clin Chem Lab Med* 2008;46:764–72.
11. Lippi G, Salvagno GL, Montagnana M, Brocco G, Guidi GC. Influence of hemolysis on routine clinical chemistry testing. *Clin Chem Lab Med* 2006;44:311–6.
12. Lippi G, Banfi G, Buttarello M, Ceriotti F, Daves M, Dolci A, et al. Recommendations for detection and management of unsuitable samples in clinical laboratories. *Clin Chem Lab Med* 2007;45:728–36.
13. Simundic AM, Topic E. Quality indicators. *Biochimica Medica* 2008;18:311–9.
14. Söderberg J, Jonsson AP, Wallin O, Grankvist K, Hultdin J. Haemolysis index – an estimate of preanalytical quality in primary health care. *Clin Chem Lab Med* 2009;47:940–4.
15. Lippi G, Salvagno GL, Blanckaert N, Giavarina D, Green S, Kitchen S, et al. Multicenter evaluation of the hemolysis index in automated clinical chemistry systems. *Clin Chem Lab Med* 2009;47:934–9.
16. Twomey PJ, Don-Wauchope AC, McCullough D. Unreliability of triglyceride measurement to predict turbidity induced interference. *J Clin Pathol* 2003;56:861–2.
17. Jeffery J, Sharma A, Ayling RM. Detection of haemolysis and reporting of potassium results in samples from neonates. *Ann Clin Biochem* 2009;46:222–5.
18. Carraro P. Potassium report of hemolyzed serum samples. *Clin Chem Lab Med* 2008;46:425.
19. Challand GS, Li P. The assessment of interpretation of test results in laboratory medicine. *Biochem Med* 2009;19:146–53.
20. Marshall WJ, Challand GS. Provision of interpretative comments on biochemical report forms. *Ann Clin Biochem* 2000;37:758–63.
21. Plebani M, Lippi G. Hemolysis index: quality indicator or criterion for sample rejection? *Clin Chem Lab Med* 2009;47:899–902.